Day 4: Informed Consumption of Evidence

How to read an academic article?

Sebastian Ramirez-Ruiz
Hertie School

Table of contents

The process of doing science

DISCOURS

DE LA METHODE

Pour bien conduire ſa raiſon,& chercher
la verité dans les ſciences.

PLUS

LA DIOPTRIQVE.

LES METEORES.

ET

LA GEOMETRIE.

Qui ſont des eſſais de cete METHODE.

A LEYDE

De l'Imprimerie de IAN MAIRE.

cIↃ IↃ c XXXVII.

Auec Priuilege.

What constitutes scientific evidence?

Scientific evidence: information gathered from *scientific research*.

Scientific research is a **systematic** investigation that aims to *discover* new knowledge, *understand* phenomena, or solve existing problems through *empirical evidence* and *logical reasoning*.

What constitutes scientific evidence?

Scientific evidence: information gathered from *scientific research*.

Hypothesis Formation

- Researchers formulate *hypotheses* based on *observations*, *prior knowledge*, or *theoretical frameworks* to guide their investigations.

Analysis and Interpretation:

- *Analyzing collected data* using statistical methods or qualitative techniques to draw meaningful *conclusions* and uncover *patterns* or *relationships*.

Data Collection

- Gathering empirical evidence through *observations*, *measurements*, *surveys*, or *experiments* to validate or refute hypotheses.

*Peer Review:

- Subjecting research findings to *scrutiny by peers* and experts in the field to *ensure validity*, *reliability*, and *adherence to scientific standards*.

What constitutes scientific evidence? (cont.)

- Through this empirical method for knowledge acquisition, researchers strive to maintain objectivity and reduce partiality (*It can be reproduced and tested*).
- The process allows researchers and research consumers to *adjust* or *discard* hypothesis based on results.
- Doing scientific research is a dynamic process, not a fixed sequence (*Not all steps occur in every inquiry, nor in the same order*).
- Scientific research builds upon existing knowledge, contributing to the *cumulative* advancement of understanding within a field through ongoing inquiry and discovery.
- Modern scientific practices increasingly emphasize *openness* and *transparency*, including data sharing, pre-registration of studies, and open access publication, promoting collaboration and accountability within the scientific community.

How does this look like?

Hertie School

***An example from my own research:**

Ramirez-Ruiz, S (2024). Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience. *Working Paper*

Politicians from 12 countries rarely engage with researchers on social media, but this can change when expertise gains salience

Sebastian Ramirez-Ruiz [1] ✉

[1]Data Science Lab, Hertie School

✉ **For correspondence:** mailto: ramirez-ruiz@hertie-school.org

Present address: Data Science Lab, Hertie School, Friedrichstraße 180, 10117 Berlin, Germany

Data availability: Data availability statement. Preprocessed data could be available on this repository.

Competing interests: The author declare no competing interests.

Abstract

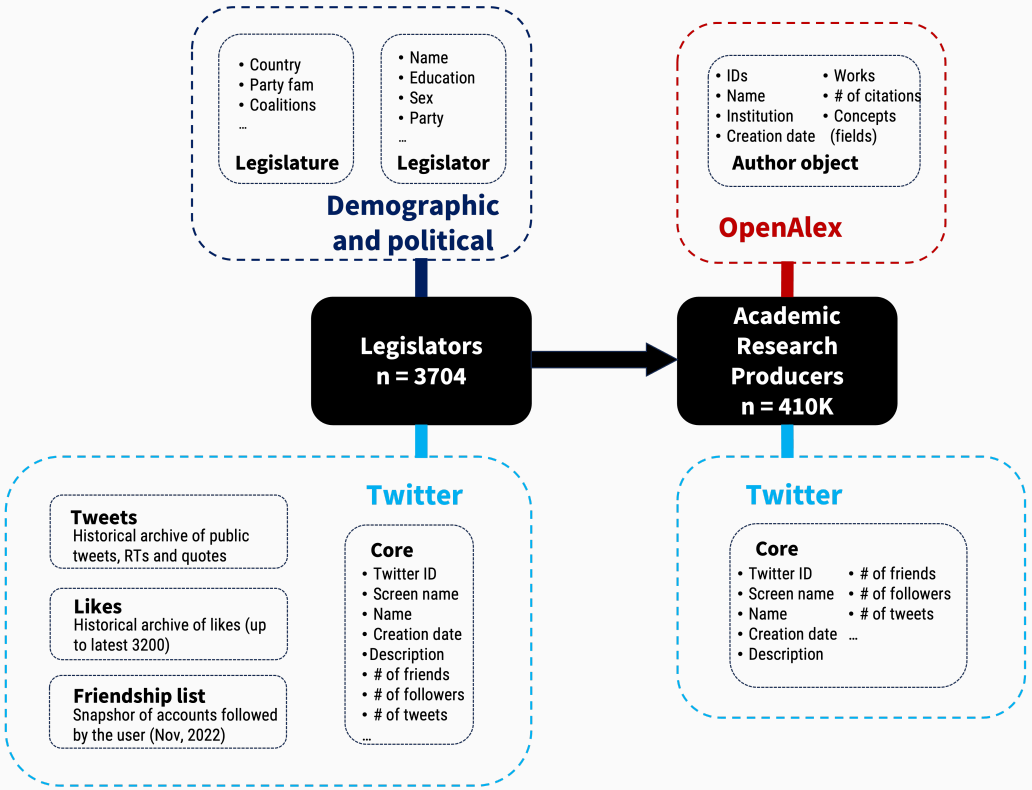
Interactions between the policy and academic communities can play an important role on political decision-making. Still, the fact that much of the policymaking process happens behind closed doors obscures our understanding of the relationships between political decision-makers with academic researchers. This paper analyzes online behavioral data from 3,704 lawmakers in 12 countries and integrates them to a novel database of 410K academic researchers on Twitter. The findings suggest that lawmakers do follow, yet rarely visibly engage with researchers online. Lawmakers from conservative and radical right parties follow and engage less with researchers online than their colleagues from other parties. While the base engagement is relatively low across legislatures, it can increase when expertise gains salience. During the early stages of the COVID-19 pandemic, marked by policy uncertainty involving a novel and technically complex policy issue, lawmakers' overall inclination to follow and engage with scholars increased, most prominently targeting researchers from the medical sciences. These findings have implications for our understanding of politicians' strategic engagement with science production.

**A sneak peek into the workflow of a Computational Social Scientist*

How does this look like?

Hertie School

a) Data structure



b) Overview of legislators

COUNTRY	SEATS	ON TWITTER	FOLLOW RESEARCHERS
Pooled	**4134**	**3704**	**86%**
Ireland	160	154	97.4%
UK	650	590	94.8%
Spain	349	314	94.3%
Colombia	108	104	92.3%
Canada	338	325	91.4%
Argentina	72	61	90.2%
Germany	736	597	89.5%
US Senate	100	100	82.3%
US House	435	431	82.2%
Italy	400	286	78.3%
Mexico	128	114	78.1%
France	577	552	73.1%
Brazil	81	76	59.2%

Were there any aspects that surprised you from the

life cycle of a scientific research project?

A primer on academic research articles

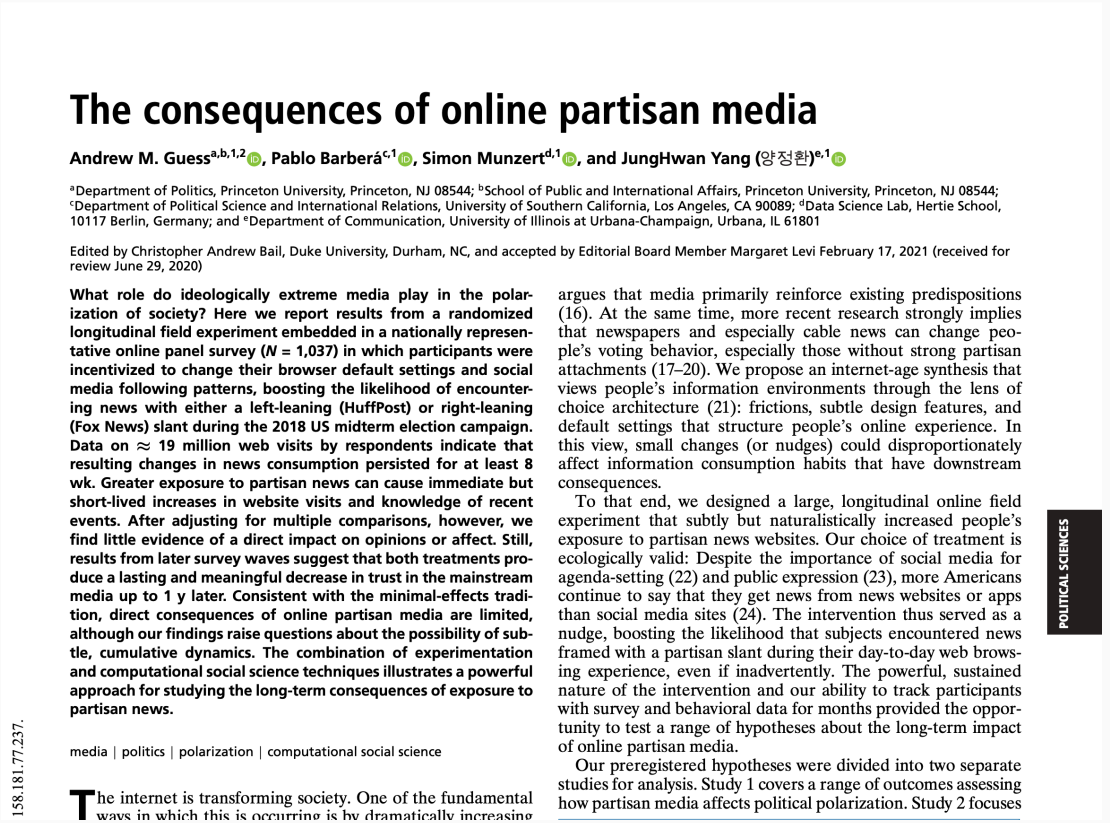
Academic research articles

Hertie School

What are they?

An **academic research article** is a **scholarly work**, mainly published in a peer-reviewed journal or proceedings. It presents original research findings, contributing to the knowledge and understanding of a specific field of study.

They are **written by experts** and generally intended for other *researchers*, *scholars*, and *professionals* within the same discipline.



The consequences of online partisan media

Andrew M. Guess[a,b,1,2], **Pablo Barberá**[c,1], **Simon Munzert**[d,1], **and JungHwan Yang (양정환)**[e,1]

[a]Department of Politics, Princeton University, Princeton, NJ 08544; [b]School of Public and International Affairs, Princeton University, Princeton, NJ 08544; [c]Department of Political Science and International Relations, University of Southern California, Los Angeles, CA 90089; [d]Data Science Lab, Hertie School, 10117 Berlin, Germany; and [e]Department of Communication, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Edited by Christopher Andrew Bail, Duke University, Durham, NC, and accepted by Editorial Board Member Margaret Levi February 17, 2021 (received for review June 29, 2020)

What role do ideologically extreme media play in the polarization of society? Here we report results from a randomized longitudinal field experiment embedded in a nationally representative online panel survey (N = 1,037) in which participants were incentivized to change their browser default settings and social media following patterns, boosting the likelihood of encountering news with either a left-leaning (HuffPost) or right-leaning (Fox News) slant during the 2018 US midterm election campaign. Data on ≈ 19 million web visits by respondents indicate that resulting changes in news consumption persisted for at least 8 wk. Greater exposure to partisan news can cause immediate but short-lived increases in website visits and knowledge of recent events. After adjusting for multiple comparisons, however, we find little evidence of a direct impact on opinions or affect. Still, results from later survey waves suggest that both treatments produce a lasting and meaningful decrease in trust in the mainstream media up to 1 y later. Consistent with the minimal-effects tradition, direct consequences of online partisan media are limited, although our findings raise questions about the possibility of subtle, cumulative dynamics. The combination of experimentation and computational social science techniques illustrates a powerful approach for studying the long-term consequences of exposure to partisan news.

media | politics | polarization | computational social science

The internet is transforming society. One of the fundamental ways in which this is occurring is by dramatically increasing

argues that media primarily reinforce existing predispositions (16). At the same time, more recent research strongly implies that newspapers and especially cable news can change people's voting behavior, especially those without strong partisan attachments (17–20). We propose an internet-age synthesis that views people's information environments through the lens of choice architecture (21): frictions, subtle design features, and default settings that structure people's online experience. In this view, small changes (or nudges) could disproportionately affect information consumption habits that have downstream consequences.

To that end, we designed a large, longitudinal online field experiment that subtly but naturalistically increased people's exposure to partisan news websites. Our choice of treatment is ecologically valid: Despite the importance of social media for agenda-setting (22) and public expression (23), more Americans continue to say that they get news from news websites or apps than social media sites (24). The intervention thus served as a nudge, boosting the likelihood that subjects encountered news framed with a partisan slant during their day-to-day web browsing experience, even if inadvertently. The powerful, sustained nature of the intervention and our ability to track participants with survey and behavioral data for months provided the opportunity to test a range of hypotheses about the long-term impact of online partisan media.

Our preregistered hypotheses were divided into two separate studies for analysis. Study 1 covers a range of outcomes assessing how partisan media affects political polarization. Study 2 focuses

POLITICAL SCIENCES

`Source`: Guess et al. (2021)

Academic research articles (cont.)

What are their building blocks?

An **academic research article** **typically follow a structured format**, including a title, abstract, introduction, literature review, methods, results, discussion, conclusion, and references.

Each **scientific domain** has their own structures and publication conventions (e.g., *political science* vs. *computer science*).

General structure of academic articles

- **Title**
  - Concise and descriptive
  - Reflects the main topic and focus of the study
- **Abstract**
  - Brief **summary** of the research (≈250 words)
- **Introduction**
  - Introduces the **research problem**
  - Provides **background and context**
  - States the research questions or hypotheses
  - Outlines the objectives and significance of the study
- **Literature Review**
  - Reviews **existing research related to the topic**
  - Identifies **gaps** or inconsistencies in the current knowledge
  - Justifies the need for the current study
- **Methods**
  - Describes the **research design and methodology**
  - Includes details on participants, materials, and procedures
  - Provides information on data collection and analysis methods

General structure of academic articles (cont.)

- **Results**
  - Presents the **findings of the study**
  - Normally uses text, tables, and figures to summarize data
  - Reports on statistical analyses and significance (in quantitative research)
- **Discussion**
  - Provides **interpretations** of the results in the context of the research questions
  - Explains the **implications** of the findings
  - **Compares** results with previous studies
  - Discusses **limitations** and suggests **directions** for future research
- **Conclusion**
  - **Summarizes** the main findings
  - **Highlights** the significance and **contributions** of the study
  - Provides closing thoughts or recommendations
- **References**
  - Lists all **sources cited** in the article
- **Appendices** (if applicable)
  - Includes **supplementary material**
  - May contain raw data, detailed descriptions of methods, or additional figures

Making sense of statistical effects

Statistical effects and effect sizes

What are statistical effects?

- Statistical effect quantify a difference or relationship between variables

Example questions about effects

- What is the marginal effect of education on income?
- How much does the probability of voting increase with age?

Statistical effect $\neq$ causal effect!

- Statistical effects are about statistical relationships between variables, not about causal relationships.
- For instance, just because your model tells you that an additional 100$ per month income are associated with 1 additional year of education, you would not conclude that income causally increases education

Effect sizes

- Effect sizes are quantitative measures of the **strength of a relationship**.
- Effect sizes express the magnitude of a difference or relationship in a standardized way.

Examples

- A (unstandardized or standardized) **group mean difference**
- The **correlation coefficient** is an effect size for the relationship between two continuous variables (see later!), e.g., $r = 0.1 \rightarrow$ weak, $r = 0.5 \rightarrow$ moderate, $r = 0.9 \rightarrow$ strong
- The **regression coefficient** expresses the predicted marginal change in an outcome relative to a unit change of the predictor (potentially conditionally on other covariates)

Example: regression effects

	Hourly wage
Education	0.505***
	(0.051)
Female	−2.275***
	(0.279)
Nonwhite	−0.119
	(0.460)
Intercept	0.650
	(0.681)
N	526
R^2	0.259
Adjusted R^2	0.255

Some advice when consuming effect sizes

Some questions to ask yourself

1. What does the effect size mean **substantively**? E.g., what does an "effect of 0.87" mean?

	Prominence	Influence
Senate	0.906***	1.483***
	(0.060)	(0.067)
Sessions served	0.163***	0.292***
	(0.016)	(0.017)
Party (Independent)	0.701*	1.059**
	(0.368)	(0.412)
Party (Republican)	0.035	−0.080
	(0.047)	(0.052)
Office: Governor	0.266*	0.450**
	(0.158)	(0.177)
Office: Lt. Governor	−0.031	0.089
	(0.257)	(0.288)
Office: US Secretary	0.551**	0.372
	(0.262)	(0.294)
Position: House Speaker	1.896***	2.670***
	(0.385)	(0.431)
Position: Majority/Minority Leader	0.185	0.711**
	(0.308)	(0.345)
Position: Whip	0.231	0.848***
	(0.233)	(0.261)
Position: Deputy Whip	0.698***	0.462*
	(0.234)	(0.262)
Position: Party Chairman	−0.115	−0.255
	(0.215)	(0.241)
(Intercept)	1.648***	1.527***
	(0.050)	(0.057)
N	492	492
R-squared	0.493	0.694
Adj. R-squared	0.481	0.687
Residual Std. Error (df = 479)	0.505	0.565
F Statistic (df = 12; 479)	38.890***	90.715***

***p < .01; **p < .05; *p < .1

Some advice when consuming effect sizes

Some questions to ask yourself

1. What does the effect size mean **substantively**? E.g., what does an "effect of 0.87" mean?

2. Is the effect size **plausible**? How does it compare to your intuition and other effects in the literature?

Political Science Research and Methods (2020), page 1 of 7
doi:10.1017/psrm.2019.63

RESEARCH NOTE

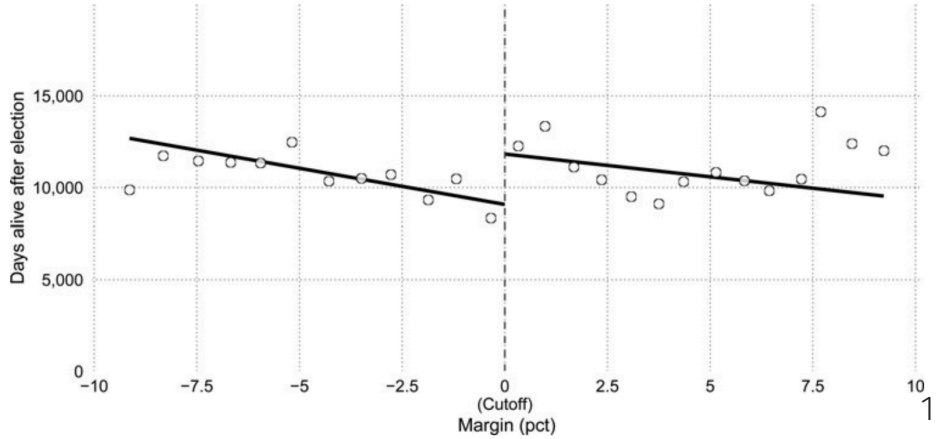
Longevity returns to political office

Sebastian Barfort[1], Robert Klemmensen[2] and Erik Gahner Larsen[3]*

[1]Independent Researcher, Copenhagen, Denmark, [2]Political Science, University of Southern Denmark, Odense, Denmark and [3]School of Politics and International Relations, University of Kent, Canterbury, Kent, United Kingdom of Great Britain and Northern Ireland
*Corresponding author. E-mail: E.G.Larsen@kent.ac.uk

(Received 18 April 2019; revised 19 September 2019; accepted 13 November 2019)

Abstract
Does political office cause worse or better longevity prospects? Two perspectives in the literature offer contradicting answers. First, increased income, social status, and political connections obtained through holding office can increase longevity. Second, increased stress and working hours associated with holding office can have detrimental effects on longevity. To provide causal evidence, we exploit a regression discontinuity design with unique data on the longevity of candidates for US gubernatorial office. The results show that politicians winning a close election live 5–10 years longer than candidates who lose.



Some advice when consuming effect sizes

Some questions to ask yourself

1. What does the effect size mean **substantively**? E.g., what does an "effect of 0.87" mean?

2. Is the effect size **plausible**? How does it compare to your intuition and other effects in the literature?

3. How **precisely** is the effect estimated?

Making sense of statistical significance

"Statistical significance" everywhere

Attractive names sustain increased vegetable intake in schools

Brian Wansink [a,*], David R. Just [b], Collin R. Payne [c], Matthew Z. Klinger [d]

[a] *Department of Applied Economics and Management at Cornell University, 15 Warren Hall, Ithaca, NY 14853-7801, USA*
[b] *Department of Applied Economics and Management at Cornell University, 16 Warren Hall, Ithaca, NY 14853-7801, USA*
[c] *New Mexico State University, College of Business, MSC 5280, PO Box 30001, Las Cruces, NM 88003-8001, USA*
[d] *Half Hollow Hills High School East, 50 Vanderbilt Parkway, Dix Hills, NY 11746, USA*

A B S T R A C T

Objective: This study will determine if the selective use of attractive names can be a sustainable, scalable means to increase the selection of vegetables in school lunchrooms.

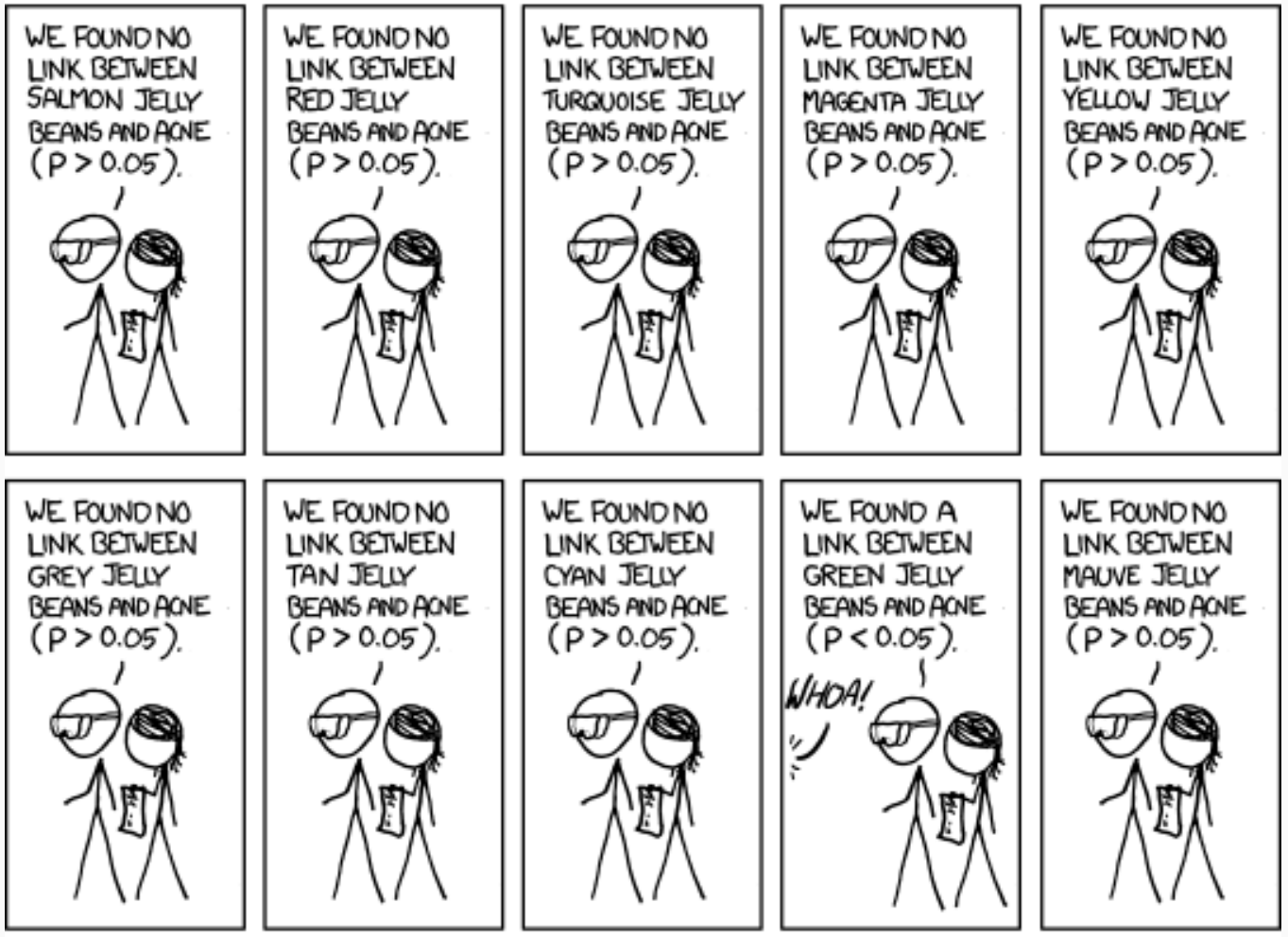
Methods: Study 1 paired an attractive name with carrots in five elementary schools (n = 147) and measured selection and consumption over a week compared to controls. Study 2 tracked food sales of vegetables in two elementary schools (n = 1017) that were systematically attractively named or not named over a two-month period. Both studies were conducted in New York in 2011.

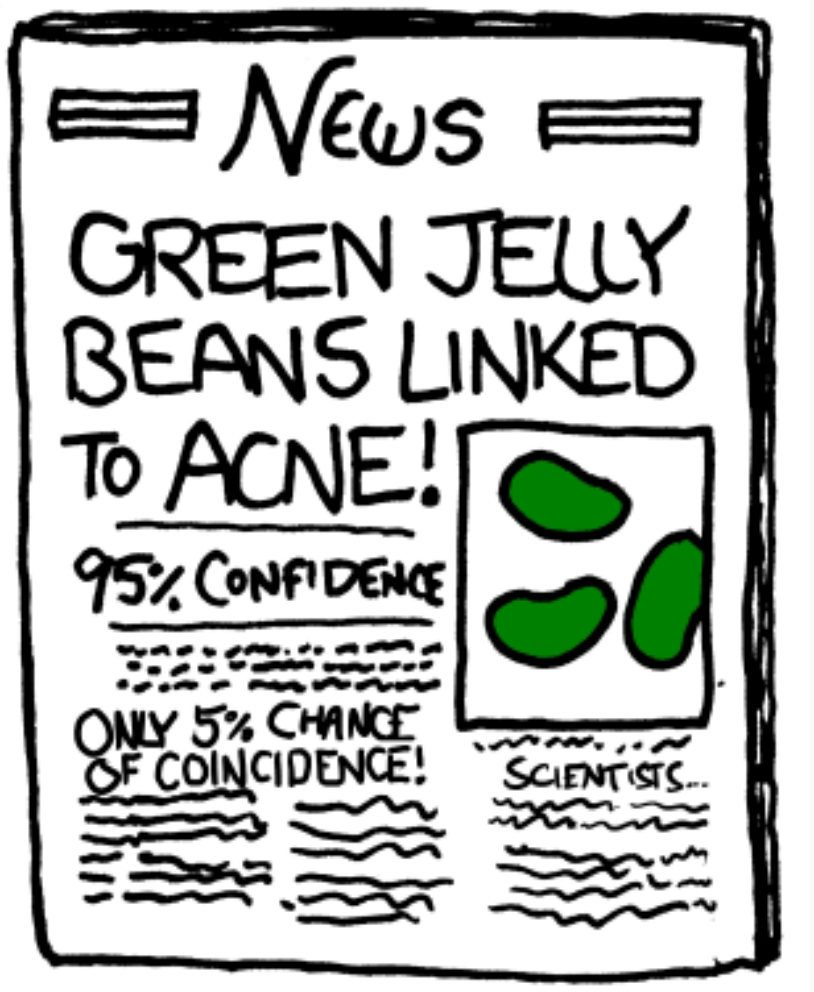
Results: Study 1 found that elementary students ate twice the percentage of their carrots if attractively named as "X-ray Vision Carrots," than if un-named or generically named as the "Food of the Day." Study 2 found that elementary school students were 16% more likely to persistently choose more hot vegetable dishes ($p<0.001$) when they were given fun or attractive names.

Discussion: Attractive names effectively and persistently increased healthy food consumption in elementary schools. The scalability of this is underscored by the success of Study 2, which was implemented and executed for negligible cost by a high school student volunteer.

`Source` Wansink et al., Retraction Watch

"Statistical significance" everywhere

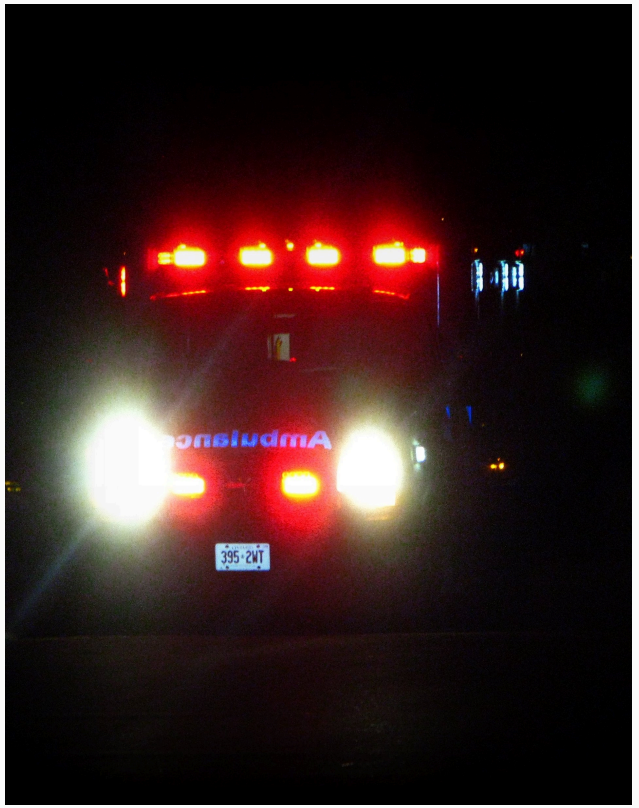




Statistical error types

A life and death example of statistical errors

You are a paramedic and you approach the scene of a car accident. One victim is laying motionless on the road and you must assess whether the victim is dead or alive, and the victim will be treated accordingly. Based on this information, **which error results in the most costly mistake?**



Hypotheses

- **Null hypothesis**: The victim is alive.
- **Alternative hypothesis**: The victim is not alive.

`Source` jeffalltogether, StackExchange.com

Statistical error types

Hypotheses

- **Null hypothesis**: The victim is alive.
- **Alternative hypothesis**: The victim is not alive.

Error types

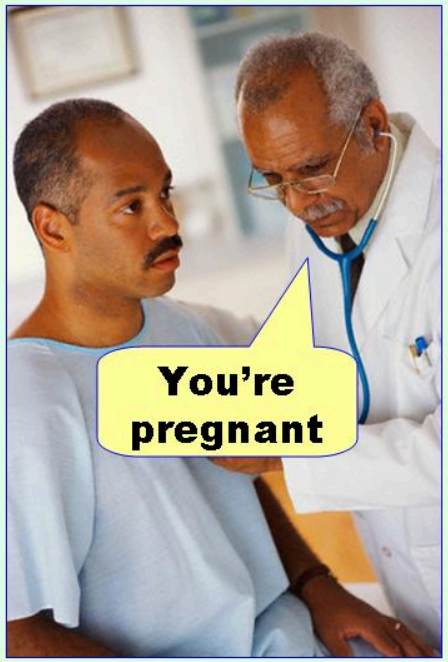
- **Type I error**: You reject the null when the null is actually true. ("false positive")
- **Type II error**: You fail to reject the null when the null is actually false. ("false negative")

Costs

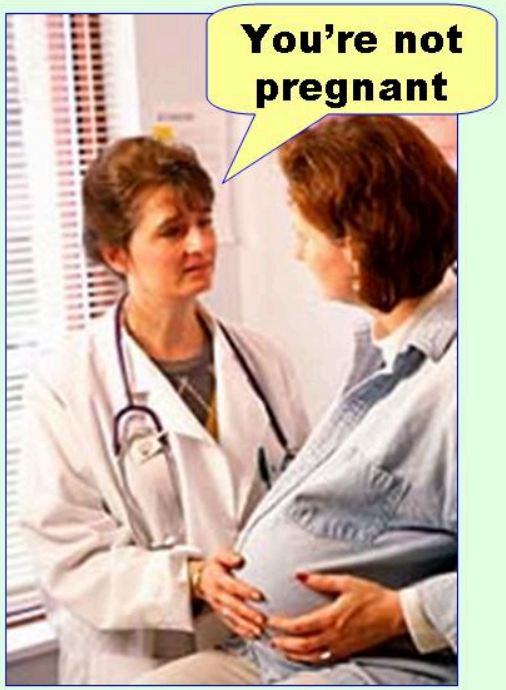
- **Type I error**: You declare the victim dead when they are actually alive. They do not receive an ambulance to the hospital for a life saving medical treatment. → **Extremely costly mistake**
- **Type II error**: You declare the victim alive when they are actually dead. You erroneously send a dead person to the hospital in an ambulance → **Not that costly mistake**

Error types illustrated

Type I error
(false positive)



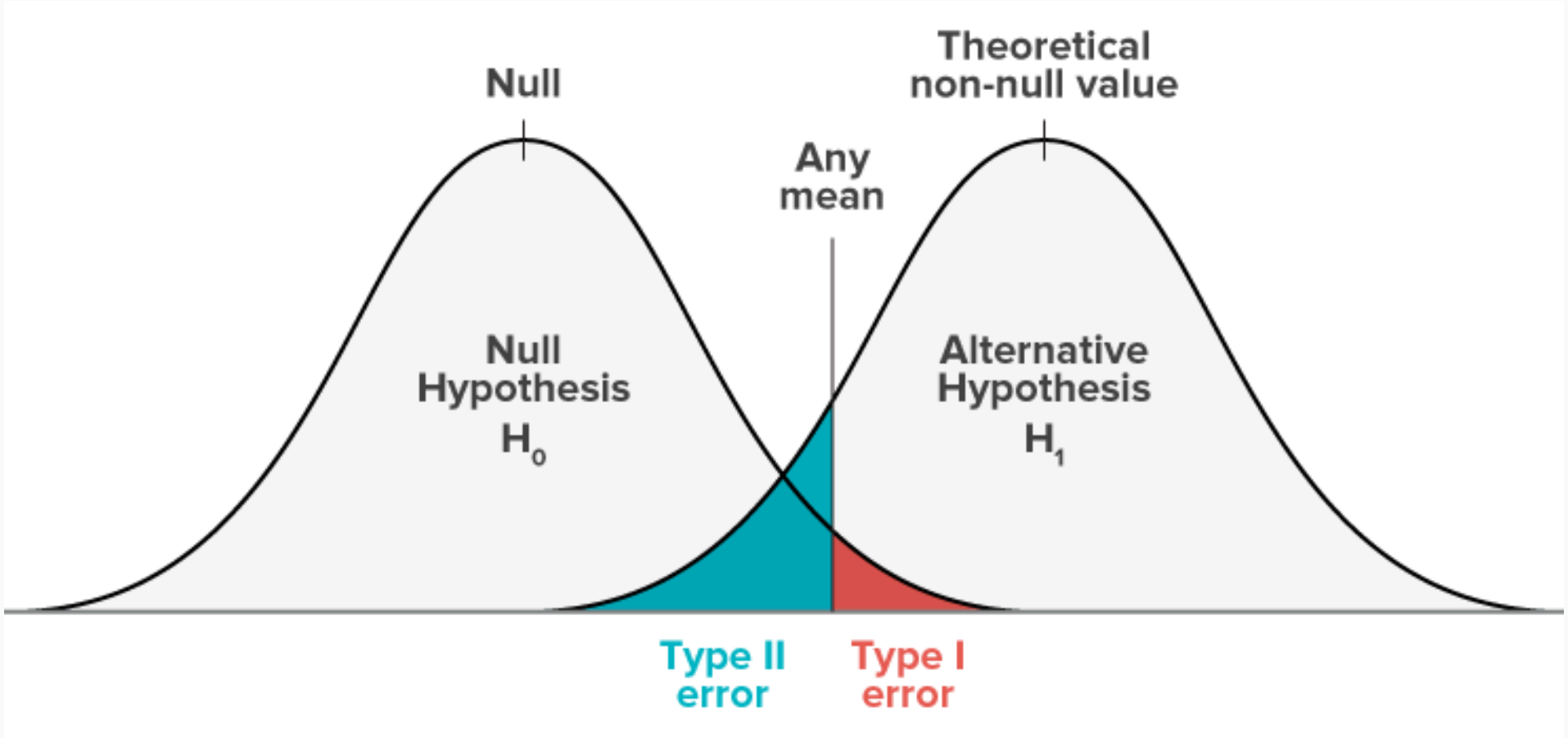
Type II error
(false negative)



Error types in hypothesis testing

		The Truth (Based on Entire Population)	
		Nothing Is There (H_0 Is True)	Something Is There (H_0 Is False)
Your Conclusion (Based on Your Sample)	I Don't See Anything (Nonsignificant)	Right!	Wrong (Type II Error)
	I See Something (Significant)	Wrong (Type I Error)	Right!

Error types in hypothesis testing



Assessing statistical significance

Statistical significance vs. practical significance

- They are not the same.
- Statistical significance is about the probability of observing the data given the null hypothesis.
- Practical significance is about the real-world importance of the result.

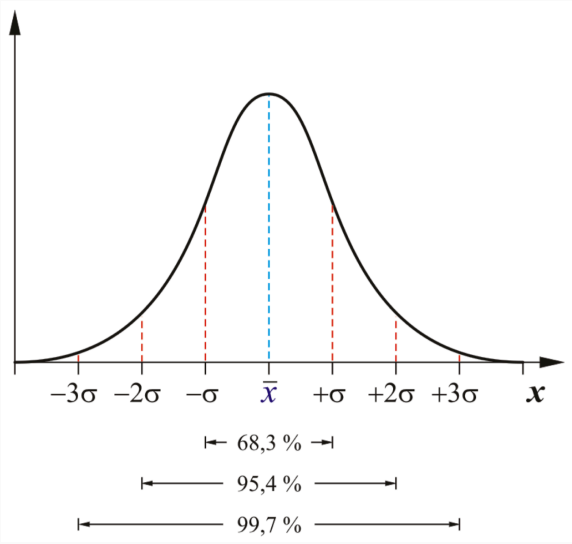
From hypothesis testing to statistical significance

Three-step approach:

1. Formulate null and alternative hypotheses.
2. Calculate a test statistic. For instance, effect size in a regression divided by standard error.
3. Compare the test statistic to a critical value; calculate a p-value.

The p-value

- The p-value is the probability of observing a result at least as extreme as the observed result if the null hypothesis were true.
- The p-value is compared to a threshold (e.g., 0.05) to decide whether to reject the null hypothesis.
- **Importantly, the p-value is not the probability that the null hypothesis is true or false!**



Eyeballing statistical significance

Hertie School

	Prominence	Influence
Senate	0.906***	1.483***
	(0.060)	(0.067)
Sessions served	0.163***	0.292***
	(0.016)	(0.017)
Party (Independent)	0.701*	1.059**
	(0.368)	(0.412)
Party (Republican)	0.035	−0.080
	(0.047)	(0.052)
Office: Governor	0.266*	0.450**
	(0.158)	(0.177)
Office: Lt. Governor	−0.031	0.089
	(0.257)	(0.288)
Office: US Secretary	0.551**	0.372
	(0.262)	(0.294)
Position: House Speaker	1.896***	2.670***
	(0.385)	(0.431)
Position: Majority/Minority Leader	0.185	0.711**
	(0.308)	(0.345)
Position: Whip	0.231	0.848***
	(0.233)	(0.261)
Position: Deputy Whip	0.698***	0.462*
	(0.234)	(0.262)
Position: Party Chairman	−0.115	−0.255
	(0.215)	(0.241)
(Intercept)	1.648***	1.527***
	(0.050)	(0.057)
N	492	492
R-squared	0.493	0.694
Adj. R-squared	0.481	0.687
Residual Std. Error (df = 479)	0.505	0.565
F Statistic (df = 12; 479)	38.890***	90.715***

***p < .01; **p < .05; *p < .1

Table 1. Summary statistics

Variable	South (1)	North (2)	Difference in means (3)	Adjusted difference in means (4)	*P* value (5)
Panel 1: Air pollution exposure at China's Disease Surveillance Points					
TSPs, μg/m^3	354.7	551.6	196.8***	199.5***	<0.001/0.002
SO_2, μg/m^3	91.2	94.5	3.4	−3.1	0.812/0.903
NO_x, μg/m^3	37.9	50.2	12.3***	−4.3	<0.001/0.468
Panel 2: Climate at the Disease Surveillance Points					
Heating degree days	2,876	6,220	3,344***	482	<0.001/0.262
Cooling degree days	2,050	1,141	−910***	−183	<0.001/0.371
Panel 3: Demographic features of China's Disease Surveillance Points					
Years of education	7.23	7.57	0.34	−0.65	0.187/0.171
Share in manufacturing	0.14	0.11	−0.03	−0.15***	0.202/0.002
Share minority	0.11	0.05	−0.05	0.04	0.132/0.443
Share urban	0.42	0.42	0.00	−0.20*	0.999/0.088
Share tap water	0.50	0.51	0.02	−0.32**	0.821/0.035
Rural, poor	0.21	0.23	0.01	−0.33*	0.879/0.09
Rural, average income	0.34	0.33	0.00	0.24	0.979/0.308
Rural, high income	0.21	0.19	−0.02	0.27	0.772/0.141
Urban site	0.24	0.25	0.01	−0.19	0.859/0.241
Predicted life expectancy	74.0	75.5	1.54***	−0.24	<0.001/0.811
Actual life expectancy	74.0	75.5	1.55	−5.04**	0.158/0.044

The sample (n = 125) is restricted to DSP locations within 150 km of an air quality monitoring station. TSP (μg/m^3) in the years 1981–2000 before the DSP period is used to calculate city-specific averages. Degree days are the deviation of each day's average temperature from 65°F, averaged over the years 1981–2000 before the DSP period. The results in column (4) are adjusted for a cubic in degrees of latitude north of the Huai River boundary. Predicted life expectancy is calculated by OLS using all of the demographic and meteorological covariates shown. All results are weighted by the population at the DSP location. One DSP location is excluded due to invalid mortality data. *Significant at 10%, **significant at 5%, ***significant at 1%. Sources: China Disease Surveillance Points (1991–2000), *China Environment Yearbook* (1981–2000), and World Meteorological Association (1980–2000).

Eyeballing statistical significance

Hertie School

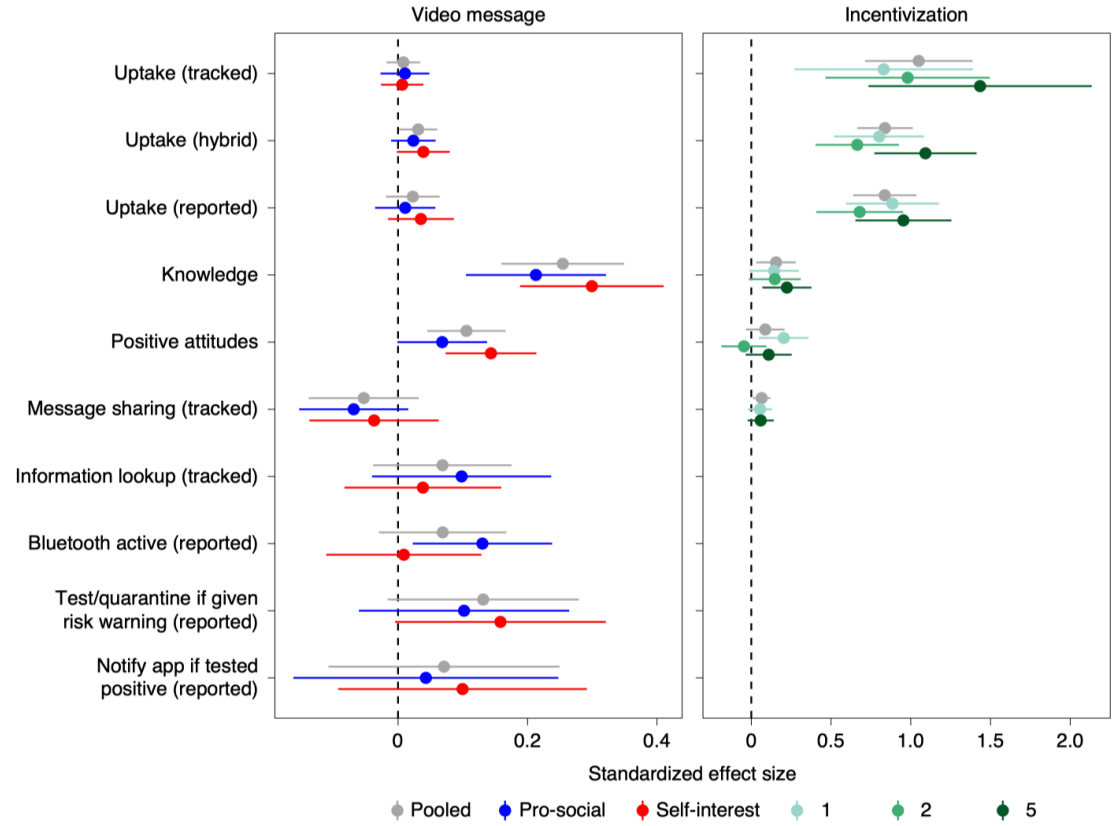


Fig. 3 | Effect of message and incentive treatments on uptake, knowledge, attitudes and behaviour. Each plot shows standardized ITT estimates with 95% CIs from fully saturated ordinary least squares regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample comprises $n = 2{,}044$, 1,356 and 1,337 respondents for estimation of the pooled, pro-social and self-interest treatment effects, respectively. The incentive sample comprises $n = 1{,}015$, 513, 516 and 494 respondents for estimation of the pooled, €1, €2 and €5 treatment effects, respectively.

Controversies around statistical significance[1]

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 2, 129–133
http://dx.doi.org/10.1080/00031305.2016.1154108

The ASA's Statement on *p*-Values: Context, Process, and Purpose

Six principles

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

[1]See also here for a nice primer to this controversy.

Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations

Sander Greenland[1] · Stephen J. Senn[2] · Kenneth J. Rothman[3] · John B. Carlin[4] · Charles Poole[5] · Steven N. Goodman[6] · Douglas G. Altman[7]

Received: 9 April 2016 / Accepted: 9 April 2016 / Published online: 21 May 2016

Abstract Misinterpretation and abuse of statistical tests, confidence intervals, and statistical power have been decried for decades, yet remain rampant. A key problem is that there are no interpretations of these concepts that are at once simple, intuitive, correct, and foolproof. Instead, correct use and interpretation of these statistics requires an attention to detail which seems to tax the patience of working scientists. This high cognitive demand has led to an epidemic of shortcut definitions and interpretations that are simply wrong, sometimes disastrously so—and yet these misinterpretations dominate much of the scientific literature. In light of this problem, we provide definitions and a discussion of basic statistics that are more general and critical than typically found in traditional introductory expositions. Our goal is to provide a resource for instructors, researchers, and consumers of statistics whose knowledge of statistical theory and technique may be limited but who wish to avoid and spot misinterpretations. We emphasize how violation of often unstated analysis protocols (such as selecting analyses for presentation based on the *P* values they produce) can lead to small *P* values even if the declared test hypothesis is correct, and can lead to large *P* values even if that hypothesis is incorrect. We then provide an explanatory list of 25 misinterpretations of *P* values, confidence intervals, and power. We conclude with guidelines for improving statistical interpretation and reporting.

Editor's note This article has been published online as supplementary material with an article of Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process and purpose. The American Statistician 2016.

Reading an academic article

A couple of things to keep in mind...

Assessing the quality of research articles is not an easy feat. There is not a one-size-fit-all device to filter the "good from the bad".

- **There are shades of quality**
  - The **good and the bad do not live in a dichotomous state**, but a continuum. Not all good (*bad*) pieces are created equally.
- **Quality ≠ Policy applicability**
  - **Not all quality studies are applicable as evidence policy**, but I argue that scientific insights employed in policy should come from high quality sources.
- **Like most things, practice makes "better"**
  - **The most you engage with academic literature** related to your policy field, **the easier it will get** to understand markers of quality.

Predatory publishing

What is it?

> Predatory publishing, also write-only publishing or deceptive publishing, is an exploitative academic publishing business model that involves charging publication fees to authors while only superficially checking articles for quality and legitimacy, and without providing editorial and publishing services that legitimate academic journals provide, whether open access or not. The rejection rate of predatory journals is low, but seldom zero. Wikipedia

Predatory publishing

What is it?

> Predatory publishing, also write-only publishing or deceptive publishing, is an exploitative academic publishing business model that involves charging publication fees to authors while only superficially checking articles for quality and legitimacy, and without providing editorial and publishing services that legitimate academic journals provide, whether open access or not. The rejection rate of predatory journals is low, but seldom zero. Wikipedia

Dr. Fraud experiment | nature | nytimes |

"An investigation finds that dozens of academic titles offered ‘Dr Fraud’ — a sham, unqualified scientist — a place on their editorial board. Katarzyna Pisanski and colleagues report."

COMMENT

EMISSIONS Don't link carbon markets, it makes for volatility p.484 | SUSTAINABILITY The invasion of Earth's largest freshwater system p.488 | EVOLUTION Four takes on whether art has a biological basis p.490 | OBITUARY Eugene Garfield, father of bibliometrics, remembered p.492

ILLUSTRATION BY RICHARD WILKINSON

Predatory journals recruit fake editor

An investigation finds that dozens of academic titles offered ‘Dr Fraud’ — a sham, unqualified scientist — a place on their editorial board.
Katarzyna Pisanski and colleagues report.

Thousands of academic journals do not aspire to quality. They exist primarily to extract fees from authors. These ‘predatory’ journals exhibit questionable marketing schemes, follow lax or non-existent peer-review procedures and fail to

journals can imitate legitimate ones that also collect fees from authors. Researchers, eager to publish (lest they perish), may submit their papers with or without verifying a journal's reputability.

established experts in the journal's field, and are considered prestigious positions.

Many predatory journals hoping to cash in seem to aggressively and indiscriminately recruit academics to build legitimate-looking editorial boards. Although academic

Some questions you can ask yourself

This is not an exhaustive list, but a helpful starting point

Research Question/Objective:

- Is the **research question clearly stated**?
- Is the **objective of the study clearly defined and aligned with the research question**?

Methodology:

- Are the **research methods** appropriate for addressing the research question?
- Is the **study design** well-suited for the research objectives?
- Are the **data collection methods** clearly described and appropriate?
- Are the (*statistical*) **analyses** appropriate for the research design and data collected?

Some questions you can ask yourself (cont.)

Validity:

- Are the **results** of the study valid, i.e., do they accurately reflect what they claim to measure? (think *internal, external, measurement, and statistical validity* with Simon)
- Are **potential biases** addressed and minimized? (think *data generation processes* with Simon) Is the methodology reliable and replicable?

Sampling:

- Is the **sampling method** clearly described and appropriate for the research question?
- Is the **sample size** adequate for the research objectives?
- Is the **sampling frame** fit? (Study vs. target populations)

Some questions you can ask yourself (cont.)

Results:

- Are the **results** clearly presented and logically organized?
- Are statistical **tests appropriately applied** and **interpreted**?

Discussion/Conclusion:

- Are the **findings discussed** in relation to existing literature?
- Are the **limitations** of the study acknowledged and discussed?
- Are the **implications** of the findings clearly stated?

Some questions you can ask yourself (cont.)

Ethical Considerations:

- Are **ethical considerations**, such as informed consent and protection of participants' rights, addressed?
- Are **conflicts of interest** disclosed?

Peer Review and Journal Reputation:

- Was the article subjected to **peer review**?
- Is the **journal/outlet reputable** and known for publishing high-quality research?

Impact and Contribution:

- Does the research make a significant **contribution** to the field?
- What is the potential **impact of the findings** on theory, practice, or policy?

Also, from **last session**...

Evaluating sources (cont.)

Many librarians encourage the *CRAAP test*[1] as a starting point to evaluate the suitability of the sources.

- **Currency**: The **timeliness** of the information.
  - When was the information published or posted?
  - Has it been revised or updated?
  - Do you need the most current information for your topic?

[1] This is a simplification of a very complex evaluation process. We will think deeper about this next session.

Evaluating sources (cont.)

Many librarians encourage the *CRAAP test*[1] as a starting point to evaluate the suitability of the sources.

- **Relevance**: The **importance** of the information for **your needs**.
  - Does the information directly relate to your topic?
  - Does it help you answer questions?
  - Who is the intended audience?

[1] This is a simplification of a very complex evaluation process. We will think deeper about this next session.

Evaluating sources (cont.)

Many librarians encourage the *CRAAP test*[1] as a starting point to evaluate the suitability of the sources.

- **Authority**: The **source** of the information.
  - Who is the author or publisher?
  - Are they qualified to write about the topic?
  - "Peer reviewed" is a good indicator for this

[1] This is a simplification of a very complex evaluation process. We will think deeper about this next session.

Evaluating sources (cont.)

Many librarians encourage the *CRAAP test*[1] as a starting point to evaluate the suitability of the sources.

- **Accuracy**: The **reliability** and correctness of the information.
  - Is the information supported by evidence?
  - Can you verify the information with another source?
  - Has the information been reviewed or refereed?
  - Does the language seem unbiased and free of emotion?
  - Can you identify any spelling, grammar or typographical errors?

[1] This is a simplification of a very complex evaluation process. We will think deeper about this next session.

Evaluating sources (cont.)

Many librarians encourage the *CRAAP test*[1] as a starting point to evaluate the suitability of the sources.

- **Purpose**: The **reason** the information **exists**.
  - What is the purpose of the information?
  - Does the point of view appear objective and impartial?
  - Is the information fact, opinion or propaganda?
  - Are there political, ideological, cultural, religious, institutional or personal biases?

[1] This is a simplification of a very complex evaluation process. We will think deeper about this next session.

Evaluating common non-academic sources in policy

Employing lateral reading, that is evaluating the credibility of a source by comparing it with other sources.

- **Verify the Source**: Determine the credibility of the organization or institution that published the document. Ask questions such as:
  - Who funds or sponsors the organization or think tank?
  - What is their reputation and track record in producing accurate and unbiased research?
  - Are there any known biases associated with the organization?
- **Check for Independent Analysis**: Look for analyses or critiques of the policy document from other reputable sources. Consider:
  - Are there other organizations or experts in the field who have reviewed or commented on the document?
  - Do they offer different perspectives or highlight any inconsistencies or shortcomings?
- **Evaluate Authorship and Expertise**: Assess the qualifications and expertise of the authors or researchers behind the document. Consider:
  - What are their credentials and affiliations?
  - Have they published other works in the field, and what is the reception of those works?
  - Are there any conflicts of interest that might influence their findings or conclusions?

Evaluating common non-academic sources in policy (cont.)

Employing lateral reading, that is evaluating the credibility of a source by comparing it with other sources.

- **Weigh against Counterarguments**: Seek outcounterarguments to the policy proposals or recommendations presented in the document. Ask yourself:
  - How do other organizations or experts interpret the same data or evidence?
  - Are there dissenting opinions within the academic or policy community?
  - Do alternative analyses provide a more comprehensive understanding of the issue?
- **Cross-Reference with Established Facts**: Verify any factual claims or statistics cited in the document by consulting reliable sources or databases. Consider:
  - Are the data sources cited in the document reputable and up-to-date?
  - Do the findings align with established research or empirical evidence?
  - Have fact-checking organizations or experts reviewed the accuracy of the information?

A checklist for policy supporting research

Hertie School

Characteristic to be checked	Quality dimensions	Most likely location in the report
1. Is it a research question?	knowing vs. prescribing	introduction, conclusion
2. Is the research question answerable?	suitability for empirical research	introduction, conclusion, executive summary
3. What kind of knowledge is needed?	data produced by design meets data required by question (descriptive, exploratory, confirmatory)	introduction, methods, conclusion
4. What order of data is required?	order of data secured (real world, experience, research context) is that required to answer question inferences justified	introduction, methods, analysis
5. What level of data is required?	level of research (sub-individual individual, collective) = level of claim inferences justified	introduction
6. What quality of data is required?	appropriate design accounts for field compromises	methods, limitations
7. What methods of analysis are required?	adequately discussed appropriate	methods, analysis, limitations
8. Do the research results support the conclusions?	sound inference accounts for threats	analysis, limitations
9. Do the conclusions provide an answer to the research question?	equivalence	introduction, conclusion

6 Closing remarks: applying the checklist in practice

In this essay we have outlined a (fast and simple) protocol that can be used by policymakers to decide if they should reject a report of empirical social science research.[11] The first criterion tested if the report was structured around a researchable question, the second if the question was answerable, the third tested for the type of knowledge claim required and the fourth proposed a number of questions to test the empirical foundations of the report. Sequentially, the most efficient procedure to quickly assess a research report is by taking the following steps:

1. First check whether the research question is proper, that is aimed at gaining knowledge, not at changing reality (if not, dismiss report on the ground that it cannot involve research).
2. Then check whether it is potentially answerable or not (if not, dismiss report on the ground that it cannot be researched).
3. Then check whether or not there is a mismatch between the conclusion and the research question (if there is a mismatch, dismiss report on the ground that the research commission has not been fulfilled).
4. When there is no apparent mismatch between the conclusion and the CRQ, check if there is a mismatch between the conclusions and the empirical research findings (if so, dismiss report on the ground that the conclusions are not substantiated empirically).
5. When there is no apparent mismatch between the results and the conclusions, check for a mismatch between study design and the kind of knowledge required to answer the RQ (if there is a mismatch, dismiss report on the ground that research is badly designed).
6. If study design and kind of knowledge are compatible, check that the order and level of data used in the study match those required to formulate the conclusion (if there is a mismatch, dismiss report on the ground that it cannot substantiate claims empirically).
7. Check whether the data collecting process and its analysis is fully documented (if incomplete or unclear, dismiss report on the ground that its data cannot be trusted and/or its methods are not transparent).
8. If the report does not immediately fail on one of the counts 1–7 listed above, more careful study of the document is in order (still keeping in mind the checklist of Table 1).

Source: Kamen and Tamás, 2014, "Should I take this seriously? A simple checklist for calling bullshit on policy supporting research", Quality & Quantity

Let's take a couple of minutes to read this article

Questions?